Genome-wide re-sequencing of multidrug-resistant *Mycobacterium leprae* Airaku-3

P. Singh[1], A. Benjak[1], S. Carat[2†], M. Kai[3], P. Busso[1], C. Avanzi[1], A. Paniz-Mondolfi[4,5], C. Peter[6], K. Harshman[6], J. Rougemont[2], M. Matsuoka[3] and S. T. Cole[1]

1) *Global Health Institute,* 2) *Bioinformatics and Biostatistics Core Facility, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland,* 3) *Leprosy Research Centre, National Institute of Infectious Diseases, Tokyo, Japan,* 4) *Departments of Biochemistry and Dermatopathology, Instituto de Biomedicina, Caracas, Venezuela,* 5) *Laboratory of Microbiology, Yale School of Medicine/Tuberculosis Reference Laboratory, VA Connecticut Healthcare, New Haven, CT, USA and* 6) *Centre for Integrative Genomics, University of Lausanne, Lausanne, Switzerland*

Abstract

Genotyping and molecular characterization of drug resistance mechanisms in *Mycobacterium leprae* enables disease transmission and drug resistance trends to be monitored. In the present study, we performed genome-wide analysis of Airaku-3, a multi-drug-resistant strain with an unknown mechanism of resistance to rifampicin. We identified 12 unique non-synonymous single-nucleotide polymorphisms (SNPs) including two in the transporter-encoding *ctpC* and *ctpI* genes. In addition, two SNPs were found that improve the resolution of SNP-based genotyping, particularly for Venezuelan and South East Asian strains of *M. leprae*.

Keywords: Leprosy, molecular epidemiology, *Mycobacterium leprae*, rifampicin-resistance, single nucleotide polymorphism
Original Submission: 29 January 2014; **Accepted:** 20 February 2014
Editor: D. Raoult
Article published online: 25 February 2014
Clin Microbiol Infect 2014; **20**: O619–O622
10.1111/1469-0691.12609

Corresponding author: S. T. Cole, Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland CH 1015
E-mail: **stewart.cole@epfl.ch**

Andrej Benjak and Solenne Carat contributed equally to this work.
†Current address: Service de Génétique Médicale, Laboratoire de Génétique moléculaire, 44093 Nantes, Cedex 01, France.

Introduction

Despite a massive decline in leprosy prevalence in the last two decades, more than 200 000 new cases of leprosy are recorded each year globally, indicating active transmission of the infection [1]. Drug resistance to one or more anti-leprosy drugs has been reported but is rare. As the causative agent, *Mycobacterium leprae*, remains uncultivable, molecular drug susceptibility testing offers a practical alternative. This involves PCR-sequencing of the drug-resistance-determining regions of the *rpoB*, *folP1* and *gyrA* genes associated with resistance to rifampicin, dapsone and ofloxacin, respectively [2]. However, a previous study described low-levels of rifampicin resistance in *M. leprae* Airaku-3, isolated from a Japanese patient who relapsed after multidrug therapy [3]. This strain exhibited phenotypic resistance to rifampicin and dapsone in a mouse footpad assay. Although the dapsone-resistance was attributable to the known *folP1* mutation (Thr53Ile), Airaku-3 has a wild-type *rpoB* sequence; hence, an explanation for its rifampicin resistance is not available [2].

Alternative mechanisms of rifampicin resistance have been described in various bacterial species, for example, the *rox* gene mediated mono-oxygenation of rifampicin [4] and duplication of the *rpoB* gene as *rpoB2* in *Nocardia* [5]. There is no orthologue of *rox* in *Mycobacterium tuberculosis* and *M. leprae*. Over-expression of RNA polymerase-binding protein A causes low level rifampicin resistance in *Streptomyces coelicolor* [6]. Its *M. leprae* orthologue *ML1439* and other relevant genes *sigA* (*rpoT*) and *sigE* were all found to be wild-type in our PCR-based analysis. As per the *M. leprae* single nucleotide polymorphism (SNP) -genotyping scheme [7], Airaku-3 belongs to SNP subtype 1D, which is the predominant genotype in many countries like Bangladesh, India, Nepal, Madagascar, Malawi and the French West Indies [7–9] and has a significant representation in Japan [7], Yemen, Venezuela [10] and the USA [11]. However, the genome of only one strain (S11-Inde2 from India) of this subtype has been sequenced [12]. Therefore, the Airaku-3 strain was selected for genome-wide sequencing using the Illumina platform to investigate the genetic basis for resistance to rifampicin and to improve the resolution of the existing SNP-genotyping scheme for *M. leprae*.

In our present study, a bacillary suspension containing c.$10E^{+08}$ cells of *M. leprae* Airaku-3 [2] in 0.1 M NaOH was passed through a 1-mL insulin syringe (0.30-mm needle) three to five times, before DNA extraction [13]. Library preparation was by the TruSeq ChipSeq method (Illumina). Paired-end sequencing, with 101 cycles on a HiSeq2000 instrument provided a total of 174.2 million reads of which 25% could

be aligned with the *M. leprae* TN genome [14], yielding >1000 times coverage, using Bowtie [15]. *De novo* assembly of these reads was performed using Velvet [16] with *k*-mer size of 91. SNPs were identified as described previously [7,12] by comparison with the other *M. leprae* strains using PCR sequencing or genome comparisons [7,11,12] to identify the unique SNPs of the Airaku-3 genome (see Supporting information, Tables S1, S2). The phylogenetic trees were obtained with SplitsTree v.4.13.1 [17] using uncorrected *p* distances, the Neighbour Joining method and 1000 times bootstrapping (Fig. 1).

After *de novo* assembly of the reads the genome of *M. leprae* Airaku-3 comprised 114 contigs ($n50$ = 42 Kb) covering 98.85% of the TN genome. All the gaps corresponded to dispersed repeats. There was no evidence of structural variations compared with the *M. leprae* TN reference genome, thereby ruling out genome rearrangements or gene duplications as an explanation for rifampicin resistance in Airaku-3. Upon comparing the genome sequence of Airaku-3 with the *M. leprae* TN [14] reference genome, a total of 114 SNPs were identified. Of these, 46 SNPs were not present in any other

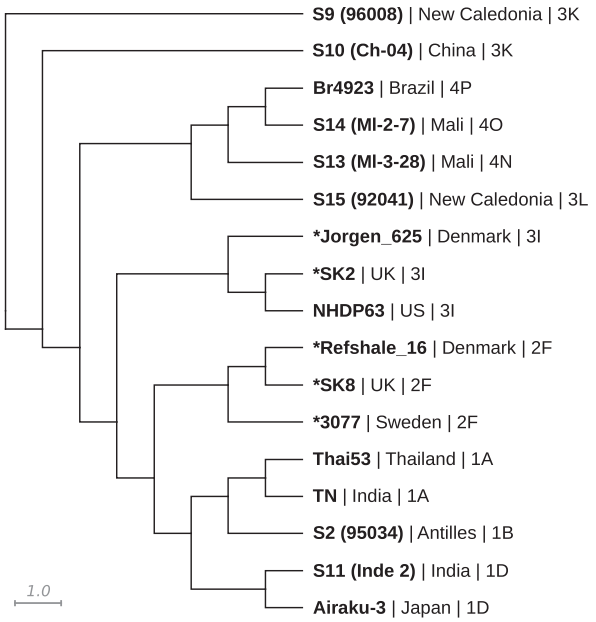

FIG. 1. Phylogenetic relationship of *Mycobacterium leprae* Airaku-3 with other *M. leprae* genomes [12]. For phylogeny, only the genomic positions in which all strains [12] had an unambiguous nucleotide call (674 positions) were considered. *Mycobacterium leprae* Airaku-3 is placed closest to another subtype 1D strain, S11-Inde2 from India, both of which share 12 SNPs that were absent in any other SNP genotypes. The geographic origins and the SNP genotypes are indicated against each strain. The ancient strains from Europe are indicated with a *. All bootstrap values are above 87 (average 99).

M. leprae genomes. These included two non-synonymous substitutions in genes with predicted transporter function belonging to the P-type ATPase family: 888973C>T in gene *ctpC* and 3209207G>A in *ctpI*. These two SNPs were also absent in the remaining *M. leprae* strains (Table S2).

Airaku-3 shared 12 SNPs with the other SNP-type 1D strain S11-Inde2 (Table S1). We analysed two of these SNPs (953582C>T and 3262657C>T) in 24 strains belonging to SNP subtype 1D from different countries (Table S2). SNP 3262657C>T correctly identified all of these 1D strains from the remaining 42 strains of other genotypes. Furthermore, the SNP 953582C>G distinguished the Venezuelan 1D strains (n = 10) from the rest of the 1D strains originating from eight different countries (Table S2). The 100 bp flanking region of this SNP in other mycobacterial outgroup species (*M. tuberculosis* complex, *M. avium* complex, *M. kansassi*, *M. marinum* and *M. ulcerans*) revealed a 'C' at the corresponding base, thereby defining the ancestral base. Therefore, Venezuelan 1D strains with a 'C' at position 953582 are designated as subtype 1D-1 while the remaining 1D strains with a derived base 'G' are termed 1D-2.

Though our present study does not reveal a clear explanation for the rifampicin resistance phenotype of Airaku-3, it has identified the unique SNPs, including two non-synonymous SNPs in transporter genes, *ctpC* and *ctpI*. However, a functional assay is required to determine whether either of these variant genes/transporters confers any degree of rifampicin resistance. Other transporters, for example, *drrA* (Rv2936), *pstB* (Rv0933) [18] and Rv1258c [19] reportedly confer low-levels of rifampicin resistance in *M. tuberculosis*. The Airaku-3 genome revealed no mutations in its *drrA* (ML2352c) whereas the orthologs of the remaining two genes (*pstB* and *ML1104c*) are pseudogenes in *M. leprae*. Rifampicin-resistant *M. tuberculosis* strains commonly possess compensatory mutations in the *rpoC* gene that restore their fitness [20]. This gene has a wild-type sequence in Airaku-3.

The comparative genomics of 1D strains in this study has discovered two useful markers: SNP3262657C>T defines the 1D genotype while the SNP953582C>G can further resolve them into 1D-1 and 1D-2. We have also identified a phylogeographic association of the 1D-1 genotype with Venezuela. These markers can be useful for molecular epidemiological studies in many countries where leprosy is endemic and the 1D genotype is predominant, and can also provide a possible explanation for disease acquisition in other countries where sporadic cases are reported in the indigenous population or among immigrants or citizens who have lived in areas where leprosy is endemic. Previously, we had successfully resolved the subtype 3I strains into 3I-1 and 3I-2 using a similar approach [11]. Hence, our present study further

exemplifies the value of genome-wide comparisons of a few strains to uncover reliable phylogeographic markers that can later be used for rapid PCR-based genotyping.

Acknowledgements

We thank the patients and clinical staff from various leprosy clinics for their participation in this study and Marc Monot and Nadine Honoré for their initial contributions in developing the SNP-typing scheme for *M. leprae*. An academic license for CodonCode Aligner software provided by Peter Richterich and Brian Campbell is also gratefully acknowledged. Part of the results of this work was presented at the 53rd Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC) meeting at Denver (6–9 September 2013, Poster presentation number C1-512). Illumina reads of *M. leprae* Airaku-3 can be accessed at the National Center for Biotechnology Information Sequence Read Archive under accession no. SRR113282.

Funding

We are grateful to the Fondation Raoul Follereau, the Brazilian–Swiss Joint Research Project (Swiss National Science Foundation) and the Global Health Institute, EPFL, Lausanne, Switzerland for financial support.

Contribution to Authorship

PS and STC designed the study with the help of MK and MM; PS, PB, CP, CA and AP-M did the sample preparation, sequencing and analysis with guidance from KH and STC; AB, SC, JR and PS performed the bioinformatic analysis; PS wrote the manuscript with input from AB, STC and other co-authors. All authors read and approved the final version of the manuscript.

Transparency Declaration

The authors declare no conflicts of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. List of all single nucleotide polymorphisms (SNPs) in *Mycobacterium leprae* Airaku-3 genome compared to the TN genome and their distribution in previously described genomes [7,12]. The unique SNPs and their effect in the Airaku-3 genome are shown in bold, with the SNPs in transporter proteins shown in blue highlight.

Table S2. *Mycobacterium leprae* isolates of various genotypes from different geographic origins, which were tested for the subtype 1D specific single nucleotide polymorphisms. The combination of 'T and C' at positions 3262657 and 953582, respectively, is characteristic of the Venezuelan 1D-1 genotype whereas a combination of 'T and G' represents the 1D-2 genotype.

References

1. World Health Organization Global leprosy: update on the 2012 situation. *Wkly Epidemiol Rec* 2013; 88: 365–380.
2. Maeda S, Matsuoka M, Nakata N, et al. Multidrug resistant *Mycobacterium leprae* from patients with leprosy. *Antimicrob Agents Chemother* 2001; 45: 3635–3639.
3. Matsuoka M. History and characteristics of isolates maintained at the leprosy research center. *Nihon Hansenbyo Gakkai Zasshi* 2010; 79: 247–256.
4. Hoshino Y, Fujii S, Shinonaga H, et al. Monooxygenation of rifampicin catalyzed by the rox gene product of *Nocardia farcinica*: structure elucidation, gene identification and role in drug resistance. *J Antibiot (Tokyo)* 2010; 63: 23–28.
5. Ishikawa J, Chiba K, Kurita H, Satoh H. Contribution of rpoB2 RNA polymerase beta subunit gene to rifampin resistance in *Nocardia* species. *Antimicrob Agents Chemother* 2006; 50: 1342–1346.
6. Newell KV, Thomas DP, Brekasis D, Paget MSB. The RNA polymerase-binding protein RbpA confers basal levels of rifampicin resistance on *Streptomyces coelicolor*. *Mol Microbiol* 2006; 60: 687–696.
7. Monot M, Honore N, Garnier T, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nat Genet* 2009; 41: 1282–1289.
8. Kuruwa S, Vissa V, Mistry N. A study of the distribution of *Mycobacterium leprae* strains among cases in a rural and urban population of Maharashtra, India. *J Clin Microbiol* 2012; 50: 1406–1411.
9. Lavania M, Jadhav RS, Turankar RP, Chaitanya VS, Singh M, Sengupta U. Single nucleotide polymorphisms typing of *Mycobacterium leprae* reveals focal transmission of leprosy in high endemic regions of India. *Clin Microbiol Infect* 2013; 19: 1058–1062.
10. Singh P, Busso P, Paniz-Mondolfi A, et al. Molecular drug susceptibility testing and genotyping of *Mycobacterium leprae* from South America. *Antimicrob Agents Chemother* 2011; 55: 2971–2973.
11. Truman RW, Singh P, Sharma R, et al. Probable zoonotic leprosy in the southern United States. *N Engl J Med* 2011; 364: 1626–1633.
12. Schuenemann VJ, Singh P, Mendum TA, et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* 2013; 341: 179–183.
13. Vansoolingen D, Hermans PWM, Dehaas PEW, Soll DR, Vanembden JDA. Occurrence and stability of Insertion Sequences in *Mycobacterium-tuberculosis* complex strains—evaluation of an Insertion Sequence-Dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J Clin Microbiol* 1991; 29: 2578–2586.

14. Cole ST, Eiglmeier K, Parkhill J, et al. Massive gene decay in the leprosy bacillus. *Nature* 2001; 409: 1007–1011.
15. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009; 10: R25.
16. Zerbino DR, Birney E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 2008; 18: 821–829.
17. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 2006; 23: 254–267.
18. Pang Y, Lu J, Wang Y, Song Y, Wang S, Zhao Y. Study of the rifampin monoresistance mechanism in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* 2013; 57: 893–900.
19. Sharma S, Kumar M, Sharma S, Nargotra A, Koul S, Khan IA. Piperine as an inhibitor of Rv1258c, a putative multidrug efflux pump of *Mycobacterium tuberculosis*. *J Antimicrob Chemother* 2010; 65: 1694–1701.
20. Comas I, Borrell S, Roetzer A, et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet* 2012; 44: 106–110.